STAT 345 Fall 2018
Homework 7 - Multiple Random Variables Name: ______________________

Please adhere to the homework rules as given in the Syllabus.

1. Consider the following joint mass function for discrete random variables X and Y.

$$f_{X,Y}(x,y) = \begin{cases} \frac{cx}{2^y}, & 1 \leq y \leq x \leq 3 \\ 0, & otherwise \end{cases}$$

The idea is that x and y can both take values 1, 2 and 3 but x must be less than or equal to y.

a) Find the value of c which makes this a valid joint mass function. *Hint: There are 6 possible combinations of X and Y. Write them out and evaluate $f(x,y)$ for each case.*

b) What is the probability that $X + Y \geq 5$?

c) Find $E(XY)$

2. Simpson's Paradox. *Simpson's paradox occurs when groups of data show one particular trend, but this trend is reversed when the groups are combined together. Understanding and identifying this paradox is important for correctly interpreting data.*

In this problem, we explore how UC Berkeley **almost** got sued for sex discrimination. In fact, statistics was used to show that there was a slight gender bias... but not in the way that we expected. The following tables summarize the admissions data from 1973. The variable X represents admission status, where $X = 1$ indicates acceptance into the program and $X = 0$ represents rejection. There are also 6 departments summarized by the random variable Y, where each value of Y corresponds to a different department.

Table 1: Admission data for Women

	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$	$Y=6$
$X=0$	0.010	0.004	0.213	0.133	0.163	0.056
$X=1$	0.049	0.009	0.110	0.071	0.051	0.130

Table 2: Admission data for Men

	$Y=1$	$Y=2$	$Y=3$	$Y=4$	$Y=5$	$Y=6$
$X=0$	0.116	0.077	0.076	0.104	0.051	0.130
$X=1$	0.190	0.131	0.045	0.051	0.020	0.008

a) Focus on Table 1 (female data). Find the marginal distributions of X and Y.

b) Focus on Table 1 (female data). Are acceptance and department (i.e. X and Y)independent? Why or why not?

c) Focus on Table 2 (male data). Find the marginal distributions of X and Y.

d) According to parts a) and c), is there a difference in marginal admission rate between men and women? Do you see why UC Berkeley was being sued?

e) Now consider only those students who applied for department $Y = 4$. Find the conditional distribution of X given that $Y = 4$. Do this separately for both men and women. Is there a strong difference between the admission rates now?

f) Repeat part e), conditional on each department $Y = 1, 2, \cdots 6$ and fill in the following table. In the end, do you think UC Berkeley committed gender discrimination?

Table 3: Conditional Probabilities

	Men	Women
$P(X = 1 \mid Y = 1)$		
$P(X = 1 \mid Y = 2)$		
$P(X = 1 \mid Y = 3)$		
$P(X = 1 \mid Y = 4)$		
$P(X = 1 \mid Y = 5)$		
$P(X = 1 \mid Y = 6)$		

3. Let $X \sim U(0,2)$, $\epsilon \sim N(0,1)$ and assume X and ϵ are independent. Now suppose $Y = 2X + \epsilon$.

a) Find the expected value and the variance of both X and Y.

b) Find $E(XY)$. *Hint: Notice that* $E(XY) = E(X(X+\epsilon))$. *Now use linearity of expectation and independence.*

c) Use the above to find $Cov(X,Y)$ and then $Cor(X,Y)$.

4. Linearity of Expectation. A mail-man must deliver mail to n different mailboxes. He has n letters to deliver, exactly one letter for each box. Unfortunately, the mail-man had a "few too many" last night and places each letter into a mailbox at random. *Note: This problem is easy. Don't overthink things.*

a) Suppose you are the owner of the i^{th} mailbox. What is the probability that you get your letter?

b) Let X_i be a Bernoulli random variable, where $X_i = 1$ if the owner of the i^{th} mailbox gets his/her letter and $X_i = 0$ otherwise. What is $E(X_i)$?

c) If we let $Y = \sum_{i=1}^{n} X_i$, then Y is the total number of people who get their letter. Using Linearity of Expectation, what is the expected number of people who get their letter?

5. Bivariate Bernoulli Distribution. A disc is checked for scratch and shock resistance. Let $X = 1$ if the disc has high scratch resistance and $X = 0$ otherwise. Let $Y = 1$ if the disc has high shock resistance and $Y = 0$ otherwise. Marginally, the distribution of X is Bernoulli with $p = 0.7$. Additionally, assume that $f_{X,Y}(1,1) = 0.6$ and $f_{X,Y}(0,0) = 0.15$.

a) Complete the table giving the Joint Probability Mass Function for X and Y.

	$x = 0$	$x = 1$
$y = 0$		
$y = 1$		

b) Give the marginal distributions of X and Y.

c) Find the correlation of X and Y.

6. Linear Functions of Random Variables.

a) Let $X_1 \sim Exp(1)$, $X_2 \sim Exp(2)$ and $X_3 \sim Exp(3)$ and assume that they are independent of eachother. Let $Y = X_1 + 4X_2 - 6X_3$. Find the mean and variance of Y.

b) Let $U \sim Binom(20, 0.5)$, $V \sim Poiss(4)$ and $Cor(U, V) = -1/3$. Find $E(W)$ and $Var(W)$ where $W = 3U - V/2$.

c) Let X_1 and X_2 be independent random variables with mean μ and variance σ^2 and let p be a constant such that $0 < p < 1$. Find the mean and variance of $Y = pX_1 + (1-p)X_2$. Use a calculus argument to show that $Var(Y)$ is minimized when $p = 1/2$, i.e. for $Y = \frac{X_1+X_2}{2}$.

7. **Challenge Problem:** Timothy throws a baseball with speed a speed of v meters per second at an angle A above the ground. The ball will land on the ground at a distance of

$$R = \frac{v^2}{g} \sin 2A$$

from where Timothy was standing (where $g = 10$ meters per second squared is the gravitational constant).

a) If $v = 10$ m/s is fixed and known, but the angle A is uniformly distributed between 0 and $\pi/2$, find the CDF and PDF of R. *Hint: Find the CDF of R first and then take the derivative.*

b) Jimothy is standing in front of Timothy. He can catch the ball if Timothy throws it between 6 and 8 meters. What is the probability that Jimothy catches the ball?

c) Now assume that the speed v is also random, and distributed as

$$v \sim Gamma(r = 1000, \lambda = 100)$$

Use Monte Carlo (notes on webpage) to approximate the mean and variance of R. Also approximate the probability that Jimothy catches the ball. Finally, provide a Histogram of the distribution of R.